

The Persuasion Hearing

Research and political reasoning



Political communication, AI persuasion and civic accountability

Created by Yoav Duman

Purpose

Use the hearing to connect political communication, causal evidence and institutional responsibility. Students must explain real comparisons, verify fictional claims, propose a positive process and rival, and defend an enforceable decision. Preserve the distinction between a capability finding, a reliability finding and a public duty. Official information and disclosed advocacy have different roles under a common factual/privacy floor. All Bridgehaven, Westmere and practice data are fictional. The exercise evaluates written reasoning about evidence and power; it does not generate campaign messages, identify a universal persuasion mechanism or demonstrate student learning gains.

Prepare the assignment

Use the simplified explanations inside the activity during class. The primary articles and further reading are optional for students. Give students time to read, ask questions and return to a passage while answering. Keep the primary sources available for your own checks.

Open the app on the teaching device, select Class and check instructor settings through 5907. Confirm which research, decision, data, transfer and frontier tasks your assignment requires.

Prepare the exact accessible charts/tables and a way to inspect source-comparison notes. Keep paper explanations available during discussion rather than expecting numerical memorization.

Keep this instructor guide and its worked answers separate from the student handout. Require commitments before later-record reveals; familiar evidence must not be described as unseen transfer.

Create a separate Canvas assignment specifying the PDF/text record and any progress JSON you want. Plan follow-up time for unfinished numerical work, the full unfamiliar transfer and final brief.

A first class session with in-app reading

This suggested first class session starts with the in-app explanations. No advance reading is required. It is a planning example, not a measured completion time; use further sessions for the rest of the inquiry.

Phase	Min	Teaching move
Opening question	5	Begin from the fictional notice and save a first idea. No advance reading is assumed.

Phase	Min	Teaching move
Learn from the explanations	15	Read the two in-app explanations together. Ask students to say in their own words what was compared, what was found and one limit. Use the vocabulary help.
First evidence inquiry	15	Write an explanation and alternative, predict a new observation, then inspect the next record. Return to the explanations when needed.
Discuss the analysis	10	Compare evidence and reasons in pairs or small groups. Resolve unclear instructions before evaluating the argument.
Save and plan continuation	5	Save progress and a backup. Continue remaining data, decision, new-case and final-record tasks in further sessions.

Continue the complete inquiry

Continue the full inquiry from the saved stopping point. Read explanations during the activity whenever needed. Keep the data, decision, new-case prediction and reasoned revision, allowing time for questions and discussion. No human student completion time has been measured.

Teaching moves

Ask for conditions before estimates. Have students attach the assigned change, comparable features, population, outcome and unit to each finding, then explain what comparing the studies changes.

Test rivalry at one link. Ask both accounts to predict the same measured outcome for the same population under one comparison. Compatible stages of a process can coexist; a well-designed policy test may leave an earlier mechanism unresolved.

Inspect one exact qualifier at a time. Require the decisive reference or missing measurement and a reason. A source link is attribution; a content error alone does not identify its contribution to attitude movement.

Make institutional feasibility visible. Count staff hours, distinguish commission from council powers and ask how a citizen obtains independent review. Treat language access as a protection in the evaluation design as well as the service.

Invite the strongest objection to each learner's own rule. Accept different feasible priorities and supported retention after counterevidence. Review reasoning rather than political agreement, grammar, initial prediction accuracy or method-help use.

Respond to misconceptions

An 81.2% increase in relative odds means 81.2% of participants were persuaded.

Article A reports a relative-odds comparison for ordered agreement. It does not report that headcount. Keep the model's unit separate from Article B's attitude points. The fictional binary parcel illustration explains ratios only; it is not a explanation of the ordinal model.

The correction proves profiles have no effect and removes the bot-versus-human finding.

The corrected G1-G0 estimate is uncertain, with an interval including no difference and positive effects. The G1-H0 finding remains. These contrasts change different features. A privacy justification can concern control over information without depending on a large added persuasion effect.

Randomized instructions prove that extra claims or falsehood caused the persuasion.

The assigned instruction package can change claim count, length, ordering and style together. Produced features do not inherit isolated causal identification. The fabrication condition added inaccuracies without a statistically clear persuasive gain; this neither proves an exact-zero effect nor removes the reliability duty.

The +8 finisher movement establishes a strong completion effect, and a larger later knowledge contrast proves the corrected guide improved learning.

Completion was selected, whereas the offer was randomized. Calculate the combined average all offered households and retain nonusers. T2 and T4 are separate fictional offer experiments without precision estimates: the larger sample knowledge contrast neither establishes a reliable improvement nor isolates first-screen coverage.

Verified facts and disclosure settle official duties, so an unusable complaint route is an acceptable experimental control.

An official guide must make eligible information and challenge rights usable; a disclosed advocate may argue within its different role. Every evaluation arm must retain complete duty-required information and usable multilingual independent complaints. Vary a lawful format/discoverability feature or use a bounded prototype with no real decision.

Debrief questions

Which positive finding from each article remains after its strongest relevant limitation, and how did it change the licensing conclusion?

Choose one audit row: what exact wording and deciding reference settle its status, and what additional record could change a not-established judgment?

Do your account and rival predict different observations for the same outcome, population and link? Which proposed processes could coexist?

What is the strongest legitimate burden of your chosen rule, who can enforce it, and what would make another feasible package preferable?

How can the finisher and all-offered means coexist? Which quantity has supplied uncertainty, and which comparison would be needed to estimate a coverage component?

What did the later Westmere record change in your condition or conclusion? Which question is now left for later, and how does every comparison group of your next test preserve usable challenge?

Assessment within six scoring homes

Score the six current criteria once each, from 0 to 3, for a maximum of 18. Use 0 for missing/materially inaccurate, 1 for emerging, 2 for developed and 3 for convincing, specific and bounded work. Factual checks and recording floors do not assign these human scores. Political agreement, grammar, support use and a correct initial guess are not criteria. Score shared source/numerical evidence in its owning criterion rather than counting it again in transfer; detailed task keys supply feedback, not extra points.

Source contribution and findings supported by the evidence

- 0 · Misstates key source contrasts or treats fictional observations as published findings.
- 1 · Names relevant findings but leaves comparison, population, unit or contribution unclear.
- 2 · Uses both sources accurately and bounds their findings; the compatible-world conclusion or consequence needs development.
- 3 · Reconstructs both procedures and units, preserves positive findings and correction uncertainty, and explains a concrete cross-study consequence.

Exact sentence verification

- 0 · Omits the audit or substitutes attribution for verification.

- 1 · Records rows but confuses a status, deciding reference, missing evidence or value judgment.
- 2 · Most classifications and reasons are correct; an exact qualifier or normative standard needs clarification.
- 3 · Defends all six exact classifications with decisive references or missing evidence, and separately argues a value standard.

Process, for the same target rival and core revision

- 0 · Treats an unobserved full chain as identified or supplies no relevant rival.
- 1 · Proposes a link but changes outcome/population between accounts or lacks differing predictions.
- 2 · Makes a bounded for the same target comparison; coexistence or the P-29 revision needs development.
- 3 · Marks observed/hypothesized links, gives genuinely differing for the same target predictions and defends a record-based core retention or revision.

Lawful feasible licensing and usable challenge

- 0 · Proposes a generic audit/ban outside the given powers, costs or rights.
- 1 · Names a plausible condition but enforcement, pass evidence, cost, burden or complaint usability is incomplete.
- 2 · Defends a feasible rule with differentiated duties; a trade-off or switch condition needs development.
- 3 · Defends a ≤ 120 -hour rule through authority, enforceable evidence, usable independent complaints, legitimate advocacy, burden and switch pattern.

Numerical comparison, population and precision

- 0 · Uses the wrong contrast, unit, population or precision claim.
- 1 · Reads values but leaves weights, offer/completion targets, representation or interval interpretation incomplete.
- 2 · Calculates correctly and distinguishes uncertainty/descriptive quantities; a sharper precision or design limit needs development.
- 3 · Shows exact contrasts and weighted calculation of the combined average, keeps units/effects you want to estimate distinct, and defends direct representation, threshold and separate-experiment limits.

Unfamiliar transfer and protected study design

- 0 · Repeats labels, ignores the later record or withholds a known right in the proposed control.
- 1 · Uses the fresh institution but leaves the revision, test target, follow-up or protection floor incomplete.
- 2 · Gives a relevant revision and lawful test; single feature or whole package scope or left for later uncertainty needs development.
- 3 · Applies the new mandate and later record, designs a comparison with clear limits with protected multilingual access and follow-up, and names left for later questions.

Worked data and case interpretation

Instructor reference. Keep these calculations and case interpretations separate from learner commitments until the corresponding work has been attempted. All local case data and added teaching examples are fictional; article findings retain their stated sources, units and uncertainty.

Article A · corrected source contrasts

Salvi's 3 September 2026 correction reports G1 versus G0, personalized versus nonpersonalized GPT-4, as +48.7% relative odds, 95% CI [−2.9%, +127.6%], $P = .0678$. G1 versus unpersonalized human H0 remains +81.2% relative odds, CI [+26.0%, +160.7%]. The direct component is uncertain, not exactly zero. These are ordered agreement outcomes among US adults in controlled debates, not percentages persuaded or votes.

Article B · its own treatment and unit

Hackenburg reports pooled added personalization of +0.43 points on its 0–100 attitude metric, 95% CI [.22, .64], compared with an 8.5-point unpersonalized persuasive effect. Its UK adult interventions differ from A's demographic-profile treatment. Preserve the positive personalization/package findings while distinguishing them from an not separately tested claim-count or error pathway. Do not subtract attitude points from relative-odds percentages.

Optional binary-odds support · fictional parcels

First group: 25/100 collect, probability 25%, odds $25/75 = 1/3$. Second group: 40/100 collect, probability 40%, odds $40/60 = 2/3$. Odds ratio = $(2/3)/(1/3) = 2$, a 100% odds increase; probability rises 15 percentage points, and 40% collect in group two. This fictional binary illustration does not explain Salvi's ordinal model or convert its headline to a headcount.

Fictional workshop · direct component and policy screen

Profiled advocacy minus neutral explanation is +6 attitude points, CI [2, 10]; unprofiled advocacy minus neutral is +5, CI [1, 9]. The supplied direct profile contrast is +1, CI [−2, +4]. Subtracting 6−5 gives the point difference but not its uncertainty. The stipulated entire-interval-above-+3 screen fails because $-2 < 3$; effects above +3 remain within the interval. Failure is neither exact zero nor equivalence. These are invented estimates under the exercise's assumptions.

Westmere T2-T3 · all-assigned outcomes and weighted reach

Each original arm contains 500 households. Matched baselines permit follow-up contrasts: support $51-50 = +1$ on the 0-100 scale; knowledge $6.6-6.4 = +0.2$ on the 0-10 test. Start rate $180/500 = 36\%$; finish rate $100/500 = 20\%$. Nonfinishers are 320 nonstarters plus 80 partial users. Calculation of the combined average: $[100 \times 8 + 400 \times (-0.75)]/500 = (800-300)/500 = +1$. Completion was not assigned. Before invitations, mean interest on a 1-5 scale was 4.2 for finishers and 2.1 for the other offered households. No precision or causal completion estimate is supplied.

Westmere T4 · a separate experiment and usable rights

The new offer comparison describes +1 support point, +0.4 knowledge points and 20% completion. Compared with T2, the knowledge sample contrast is 0.2 points larger, but neither contrast supplies precision and the trials do not isolate coverage. Among offered households using a nonmajority language, half report that the complaint form opens only in the majority language. The renewal bonus/version ranking identifies an incentive, not its causal contribution. Require complete official information and usable multilingual independent complaints in every test arm; no application, ballot or long-term outcome is measured.

Independent fictional library representation

The added-call contrast is text plus phone call minus text: -0.3 mean visits per assigned household over four weeks, CI [−0.9, +0.3]. Text minus postcard is +0.8 [0.2, 1.4]; text plus call minus postcard is +0.5 [−0.1, 1.1]. Plot the direct component in mean-visit differences with zero and its exact table. Although $0.5-0.8 = -0.3$, subtracting separate interval endpoints cannot derive component uncertainty. These are invented values, not attendance percentages.

Fictional licensing budget and powers

Package A: $80 + 40 = 120$ hours. Package B: $100 + 20 = 120$. Package C: $60 + 40 + 20 = 120$. A costed adaptation may also qualify. The commission can inspect, require five-working-day corrections, pause the public contract and publish reasons; the independent clerk can order a service correction. The council controls the route and public spending. Staff-hour costs promise no treatment effect, and a public-contract pause does not prohibit lawful independent advocacy.

Review the workspace products

Audit the six exact factual sentences

Key: (1) supported by T's weekday 06:00 start; (2) contradicted by weekend 08:00; (3) not established because B's 1.2-million ceiling is not a 900,000 operating-cost estimate; (4) supported by F's standard adult 2-credit fare; (5) contradicted by the eligible registered low-income 1-credit fare; (6) not established without a representative preference poll. Inspect every reference reason, not only status matches. "Fairest" requires a defended value standard and comparison.

Offer, reach and completion evidence

Require both T2 contrasts, their different score scales, the original 500-per-arm denominators and the exact weighted calculation of the combined average shown above. The offer effect you want to estimate retains nonusers; +8 among selected finishers does not identify completion's effect. Inspect the proposed use/completion design's target and assignment/compliance/selection assumptions. Prior-interest adjustment alone is insufficient, and voluntary access must remain protected.

Revised inference and a test of one link

Check that workshop intervals remain distinct from T2/T4 descriptive contrasts and that the two offer experiments are not treated as a randomized coverage contrast. The account and rival must share outcome, population and chosen link, with meaningfully differing predictions. Inspect what is assigned and held comparable, single feature or whole package scope, nonuser and missing-follow-up treatment, usable multilingual independent complaints in every comparison group and the earlier question now left for later. A policy-access test can be valuable without identifying the source of support movement.

All seven evidence check keys

What did the correction change?

Check comparison. Which statement uses both comparisons correctly?

Supported option B: The bot-versus-human finding remains, while the direct personalization comparison leaves uncertainty about what profiles add to GPT-4.

This preserves the supported opponent comparison and the corrected uncertainty about personalization. An uncertain estimate also does not demonstrate that profiles have no effect.

Rationale to inspect: Use A's two named contrasts to explain your choice. State what differs, what remains, the measured unit and the finding the correction does or does not change.

Salvi et al., corrected Aggregate results, and Author Correction, 3 September 2026.

Influence rose. Why did accuracy fall?

Check scope. Which interpretation respects both findings?

Supported option C: The results establish that persuasion and mistakes can rise together without establishing that the mistakes caused the persuasion.

This distinguishes the observed trade-off from its cause. It preserves the reliability concern without inventing a demonstrated causal advantage from false claims.

Hackenburg et al., How accurate is the information provided by the models?, published p. 6.

An offer and a selected group

Check transfer. Which conclusion uses the assignment and populations correctly?

Supported option B: The randomized offer contrast is +0.5 recall points; the selected finisher mean answers a different question and does not itself identify a completion effect.

The offer comparison retains all assigned visitors. Finisher movement can reflect conversation and earlier differences; it is not the randomized offer effect.

Invented different-topic practice; the required Westmere transfer uses its own records.

Effects and institutional duties

Check ruling. Which recommendation follows from the stated evidence and duties without substituting one outcome for another?

Supported option D: Keep privacy, encountered information and measured outcomes distinct; apply official coverage/challenge duties and disclosed-advocacy duties through the institution's mandate.

The studies constrain effects and uncertainty; the charter supplies different institutional duties. Verification, profile inspection, encountered alternatives and citizen understanding require distinct evidence.

Rationale to inspect: Name one official-service duty and one permitted advocacy role. Explain which supplied evidence measures each, and what additional pass evidence your recommendation needs.

Application to an invented charter; no published evaluation of these licensing conditions.

Use the direct contrast

Check data-read. Which reading of the profile contrast is justified?

Supported option B: The estimate is +1 score point, with an interval from -2 to $+4$; the data do not establish a clear positive profile advantage.

The direct contrast estimates the added profile difference. Its interval includes zero and negative as well as positive values, so it does not establish a clear positive advantage.

The display is a teaching exercise, not an estimate from either published paper.

Apply a stated evidence threshold

Check data-infer. For this exercise, the commission's evidence screen requires the entire 95% interval for the added profile effect to lie above +3 attitude-score points. Which assessment follows?

Supported option A: The screen is not met: the direct profile interval, -2 to $+4$, extends below +3.

The direct added-profile interval does not lie entirely above +3. Values above +3 remain within it, so failing this screen does not prove that an effect of that size is impossible or that the effect is exactly zero. The screen is a stipulated policy rule, not a scientific requirement to choose a particular safeguard.

Rationale to inspect: Identify the direct added-profile interval and apply the stated whole-interval rule. Explain why the assessment differs from an equivalence or impossibility claim.

The display is a teaching exercise, not an estimate from either published paper.

Which display answers the added-call question?

Check data-design. Which display preserves the target comparison, axis and uncertainty?

Supported option C: Show the text-plus-call minus text row on a mean-visit-difference axis with zero, its supplied interval and the exact accessible table.

This keeps the text reminder as the reference, plots the directly estimated added component in mean visits, and preserves its own uncertainty. The -0.3 point is not a household participation rate or a claim of exact zero.

Rationale to inspect: Explain why your chosen contrast and axis answer the added-call question. Use the named conditions, outcome unit and supplied uncertainty; state the precise representation mistake avoided.

Invented independent representation exercise; no article estimate or measured learning claim.

Instructor controls and the 5907 panel

Open Instructor controls and enter 5907 for this activity. This opens the local workbench, including task keys and check answers. To manage online classes, connect the saving service and enter 5907 again in the server workbench. Canvas remains a separate login and assignment.

Set the default Class or Fun route and which routes are available. Assignment switches cover research notes, political decision, data, unfamiliar transfer, structured workspaces, written justifications, the representation check and optional frontier availability or requirement. Select full or core response guidance; choose unlimited retries or first recorded answers, and feedback after submission or on request. Method hints, unrelated worked examples, badges and deliberate note reuse can also be assigned.

Save and export the assignment settings, or import a compatible settings file and inspect the staged changes before saving. Learner preferences include reading size, high contrast, reduced motion, art visibility and guided responses. Settings, learner work and display preferences are separate records.

Review the six human criteria and task key separately from completion. Load the learner record or an explicitly submitted snapshot to record a score and evidence for each criterion; do not treat word floors, mechanical checks or supported factual choices as a semantic grade.

The private instructor recovery key is separate from 5907; use it to reopen your owned classes. Instructor sessions and class ownership remain isolated. The server workbench can save defaults, create owned classes, update assignment switches and view explicitly submitted snapshots. Private personal drafts are not silently turned into submissions. Keep local progress and settings backups as well.

Private drafts consent and submission

Online saving is optional. Connecting to a saving service does not upload or submit the device draft. Use Save online personal draft only when you want to create or update that personal copy; keep its private recovery key and your local JSON backup.

A personal online draft is not an instructor submission. To submit through a configured hosted class, find the class with its student submission code, review the destination, apply the assignment switches if requested, complete the Class record and save the current version online. Tick the explicit consent box and use Submit this response snapshot.

Submission sends that version to the identified class. Later edits stay personal until you explicitly submit again. This hosted-class route is separate from the Canvas assignment and its upload confirmation.

Device backups and Canvas

Keep compatible progress JSON and settings backups. A readable TXT or PDF export supports review; the JSON preserves continued work. Saving in the player, saving a personal online draft and explicitly submitting to a hosted class are different actions. Canvas remains a separate assignment upload; verify its requested format and submission receipt.

Research sources

Francesco Salvi, Manoel Horta Ribeiro, Riccardo Gallotti, and Robert West (2025 · corrected 2026). On the conversational persuasiveness of GPT-4. *Nature Human Behaviour*. 9: 1645–1653 · DOI 10.1038/s41562-025-02194-6

[Complete primary text](#)

Kobi Hackenburg, Ben M. Tappin, Luke Hewitt, Ed Saunders, Sid Black, Hause Lin, Catherine Fist, Helen Margetts, David G. Rand, and Christopher Summerfield (2025). The levers of political persuasion with conversational artificial intelligence. *Science*. 390(6777): eaea3884 · DOI 10.1126/science.aea3884

[Complete primary text](#)